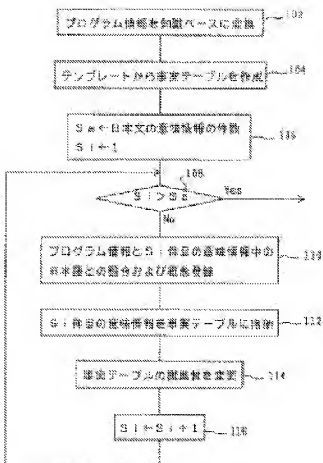


KNOWLEDGE BASE GENERATING METHOD

Inventor(s): MIYAMOTO HIRONORI; KONDO HIDEFUMI ±
Applicant(s): HITACHI LTD ±
Classification: - international: **G06F17/28; G06F9/44;** (IPC1-7): G06F15/38; G06F9/44
 - European:
Application number: JP19920302059 19921112
Priority number(s): JP19920302059 19921112

Abstract of JP6149574 (A)

PURPOSE:To determine vocabulary of the language used in a computer program with respect to words to be stored in a knowledge base which do not exist in a dictionary but can be recognized as words used in the computer program or words having meanings equivalent to those in respect of the semantic structure of a Japanese-language sentence. **CONSTITUTION:**A list of names is converted to a knowledge base (102), and case information of the Japanese language is converted to a table structure expressing facts and a concept in the higher order of the concept stored in a column of the table structure correspondingly to the case. When a word included in semantic information of a Japanese-language simple sentence is semantically equivalent to a name used in the computer program described in the computer language, the word included in semantic information is registered as a synonym; but when it is not



semantically equivalent, the word included in semantic information is registered as a new concept in the lower order of the concept corresponding to the case (110), and semantic information of the Japanese-language simple sentence is stored in the table structure expressing facts (112), and higher-order concepts in respective columns of the table structure expressing facts are changed in accordance with classifications of concepts stored in the table structure expressing facts.

Last updated: 14.03.2012 Worldwide Database 5.7.38; 93p

特開平6-149574

(43)公開日 平成 6年(1994) 5月27日

(51)Int.Cl. ⁵	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 9/44	3 3 0 P	9193-5B		
15/38	M	7323-5L		

審査請求 未請求 請求項の数 5 (全 24 頁)

(21)出願番号	特願平4-302059	(71)出願人	000005108 株式会社日立製作所 東京都千代田区神田駿河台四丁目 6 番地
(22)出願日	平成 4年(1992)11月12日	(72)発明者	宮本 博紀 神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内
		(72)発明者	近藤 秀文 神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内
		(74)代理人	弁理士 小川 勝男

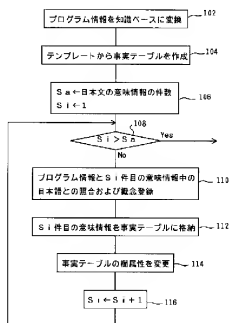
(54)【発明の名称】 知識ベース生成方法

(57)【要約】

【構成】名称の一覧を知識ベースに変換し(102)、日本語の格情報を事実を表す表構造と格に対応して表構造の欄に格納する概念の上位概念に変換し、日本語の単文の意味情報に含まれる語が計算機言語で記述されたコンピュータプログラム中で使用される名称と意味的に等しければ意味情報に含まれる語を同義語として登録し、等しくなければ意味情報に含まれる語を格に対応する概念の新規の下位概念として登録し(110)、日本語の単文の意味情報を事実を表す表構造に格納し(112)、事実を表す表構造に格納した概念の種類に応じて事実を表す表構造の各欄の上位概念を変更する(114)。

【効果】知識ベースに格納する語が辞書にない語でも、コンピュータプログラム中で使用される言葉、または、これと等価な意味を持つ言葉であることが日本語の意味構造からいえれば、知識ベースに格納する語はコンピュータプログラム中で使用される言葉の語彙が決定できる。

図 1



【特許請求の範囲】

【請求項1】概念の階層構造と、日本語の格に対応する欄を持ち、前記欄ごとに概念の種類が決まっている事実を表す表構造となる知識ベースを生成する方法であって、計算機言語で記述されたコンピュータプログラム中で使用される名称の一覧を前記知識ベースに変換し、日本語の格情報を実事を表す表構造と格によって決まる種類の概念に変換し、日本語の単文の意味情報に含まれる語が前記計算機言語で記述された前記コンピュータプログラム中で使用される名称と意味的に等しければ意味情報に含まれる語を一致した概念の同義語として登録し、等しくなければ意味情報に含まれる語を格によって決まる種類の概念の新規の下位概念として登録し、日本語の単文の意味情報を実事を表す表構造に格納し、事実を表す表構造に格納した概念の種類に応じて事実を表す表構造の各欄の概念の種類を変更すること特徴する知識ベース生成方法。

【請求項2】請求項1において、ある処理を記述したプログラム中で使用される名称の一覧を知識ベースに変換し、そのプログラムの仕様を記述した日本語の単文の意味情報を知識ベースに変換する知識ベース生成方法。

【請求項3】請求項1において、日本語の単文の意味情報から構文上等価な関係で結ばれる複数の語をとりだし、計算機言語で記述されたコンピュータプログラム中で使用される名称が複数の語のいずれか一つと同じならば、日本語の単文の意味情報と計算機言語で記述されたコンピュータプログラム中で使用される名称が意味的に等しいと判別する知識ベース生成方法。

【請求項4】請求項1において、事実を表す表構造に格納する意味情報中の語に対応する概念の一つ上位の概念を表構造の欄属性に追加することで事実を表す表構造の各欄の概念の種類の変更を行う知識ベース生成方法。

【請求項5】請求項1において、計算機言語で記述されたコンピュータプログラム中で使用される名称とそのプログラムの二次記憶装置上での位置を知識ベースに変換する知識ベース生成方法。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は知識ベースの生成方法に係り、特に、コンピュータプログラムの開発者などが使用する情報検索用の知識ベースの生成方法に関する。

【0002】

【従来の技術】日本語の文章から概念階層構造を持つ知識ベースを自動的に生成する試みは従来から行われてきたが、そのためには日本語の文章中の言葉に網羅する辞書が必要とされている。この辞書には日本語の文章中の言葉が概念階層構造中のどの階層にあてはまる言葉であるかという語彙情報が含まれる。

【0003】しかし、必ずしも全ての言葉が辞書に登録されているわけではない。これは、とくに、プログラム

仕様などから知識ベースを生成する場合にいえることである。プログラム仕様書などに含まれる言葉はプログラム仕様書の作成者が作った言葉が多く、これらは辞書に現れることは少ない。また、プログラム仕様書などにコンピュータプログラムの名称などが現れる場合もあり、このような言葉は一般に辞書には含まれないと考えられる。

【0004】

【発明が解決しようとする課題】プログラム仕様書などに書かれている日本語の文章に含まれる言葉が辞書に含まれていなければ語彙情報などがわからないため、その言葉の概念階層構造中の正確な格納位置が決定できなくなる。

【0005】本発明の目的は、辞書に含まれないため概念階層構造中の正確な位置に格納できない言葉を減少させるための方法を提供することにある。

【0006】

【課題を解決するための手段】計算機言語で記述されたコンピュータプログラム中で使用される名称の一覧を知識ベースに変換し、日本語の格情報を実事を表す表構造と格に対応して表構造の欄に格納する概念の上位概念に変換し、日本語の単文の意味情報に含まれる語が計算機言語で記述されたコンピュータプログラム中で使用される名称と意味的に等しければ意味情報に含まれる語を同義語として登録し、等しくなければ意味情報に含まれる語を格に対応する概念の新規の下位概念として登録し、日本語の単文の意味情報を実事を表す表構造に格納し、事実を表す表構造に格納した概念の種類に応じて事実を表す表構造の各欄の上位概念を変更することによって解決する。

【0007】

【作用】計算機言語で記述されたコンピュータプログラム中で使用される名称の一覧を知識ベースに変換することによってコンピュータプログラム中で使用される名称から概念階層構造を作成し、日本語の格情報を実事を表す表構造と日本語の格情報を実事を表す表構造と格によって決まる種類の概念に変換することで日本語の単文の事実を表す格納するための入れ物と表構造にこれらの概念の上位概念を用意し、日本語の単文の意味情報に含まれる語が計算機言語で記述されたコンピュータプログラム中で使用される名称と意味的に等しければ意味情報に含まれる語を同義語として登録し、等しくなければ意味情報に含まれる語を格によって決まる種類の概念の新規の下位概念として登録することで単文の意味情報中の語を既存の概念構造中の正しい位置に組み込み、日本語の単文の意味情報を実事を表す表構造に格納することによって日本語の単文を事実を表す知識として知識ベースに格納し、事実を表す表構造に格納した概念の種類に応じて事実を表す表構造の各欄の概念の種類を変更することで事実を表す表構造の欄の概念の種類を合わせる。

【0008】

【実施例】この実施例は、プログラムの名称一覧から構成されるプログラム情報と日本語の単文の意味情報を読み込み、知識ベースを生成する方法である。この実施例で特徴となるのは日本語の意味情報から等価な意味を持つ複数の語を見つけ、そのうちの一つがプログラムの名称と一致すれば、日本語の意味情報から取り出した複数の語群は一致したプログラム名称の同義語として知識ベースに格納することである。

【0009】図1のフローチャートの説明を行う前に、図2で図1の生成方法を実現する場合の構成についての説明を行い、図3、図4、図5、図6で図1で現れるデータについての説明を行い、その後、図1の説明を行う。

【0010】図2は、知識ベース生成方法を実現するブロック図で、図1で述べる処理方法から知識ベース生成プログラム220として実現される。知識ベース生成プログラム220は主記憶装置202、CPU204、ファイルシステム206、キーボード208、CRT210で構成される電子計算機システム上で動作する。

【0011】本実施例で用いるプログラム情報212、テンプレート214、日本語の意味情報216、知識ベース218、知識ベース生成プログラム220は全て論理的には主記憶装置202上に存在する（つまり、仮想記憶も含む）。

【0012】図3は、プログラム情報212の構成を表す図で、プログラム情報212はプログラムの中で使用されているサブルーチン名の一覧や外部変数の名前と型の一覧や自由変数の名前と型および自由変数の存在するサブルーチン名の一覧である。すなわちプログラム情報212は、サブルーチンテーブル302、外部変数テーブル304、自由変数テーブル306、および、それらのテーブルの一覧308からなる。

【0013】サブルーチンテーブル302は「名前」欄302aから構成されるデータ構造である。外部変数テーブル304は「名前」欄304a、「型」欄304bから構成されるデータ構造である。自由変数テーブル306は「名前」欄306a、「型」欄306b、「含む」欄306cから構成されるデータ構造である。テーブル一覧308は「テーブル名」欄308aから構成されるデータ構造である。

【0014】サブルーチンテーブル302はプログラム中のサブルーチンの一覧であり、外部変数テーブル304は外部変数の名前と型の型の一覧であり、自由変数テーブル306は自由変数の名前と型と自由変数の存在するサブルーチン名の一覧である。

【0015】図4は、テンプレート214の構成を表す図で、テンプレート214は日本語の格情報を表すテーブル形式のデータ構造である。テンプレート214は「動詞」欄214a、「格」欄214b、「属性」欄2

14cを持つ。テンプレート214は格情報として持つことのできる概念を定義している。図4の場合、例えば、動詞「行」の「は」という格は「有意味性」という概念の格情報を持つことを表す。

【0016】図5は、日本語の意味情報216の構成を表す図である。この図には2件の意味情報を記述している。日本語の意味情報216は日本語の単文の意味的な構造を表すデータ構造で、動詞テーブル502とノード504からなる。

【0017】動詞テーブル502は動詞の語502a、格名称フラグ502b、ポインタ502cおよびID欄502dを持つ。動詞の語502aは動詞の終止形であり、格名称フラグ502bは動詞の持つ格の種類を表すフラグであり、ポインタ502cは格情報となるノード504へのポインタであり、ID欄502dはポインタ502cの指すノードに対応する概念IDを入れるものである。

【0018】ノード504は語504a、品詞504b、格名称フラグ504c、ポインタ504dを持つ。語504aは格情報となる言葉であり、品詞504bは語504aの品詞であり、格名称フラグ504cは語504aの格の種類を表すフラグであり、ポインタ504dは他のノードへのポインタである。

【0019】図5では、例えば「行」という動詞の「は」格の言葉は「主プログラム」であり、「を」格の言葉は「数値計算」であり、また、「主プログラム」という言葉と意味的に等価（格名称フラグ504cに“=”という格を持つ）な言葉として「main」が存在することを表す。

【0020】図6は、知識ベース218の構成を表す説明図である。知識ベース218はネームテーブル602、階層関係604、事実テーブル606からなる。

【0021】ネームテーブル602は概念の一覧を表すデータ構造であり、「ID」欄602a、「名前」欄602b、「同義語」欄602cを持つ。「ID」欄602aは知識ベースで用いる概念の識別子を入れる。「名前」欄602bは概念の名前を入れる。「同義語」欄602cは概念の別名を入れるリスト構造へのポインタを入れる。ネームテーブルには初期状態として、ID=0、名前=「概念」が入る。

【0022】階層関係604は概念の上下関係を表すデータ構造であり、「上位」欄604a、「下位」欄604bを持つ。「上位」欄604aは概念のIDが入り、「下位」欄604bにはその下位概念のIDが入る。

【0023】事実テーブル606は事実を入れるテーブル構造である。事実テーブルは一つの知識ベースに任意個作ることができる。それぞれの事実テーブルにはテーブル名が付き、テーブルを識別するときはテーブル名を用いる。

【0024】図6の事実テーブル606では事実テーブルのデータ形式について述べる。

【0025】事実テーブル606は1個以上の欄を持つことができる。欄は格フラグ606a、属性ポインタ606b、属性リスト606c、リストポインタ606dで構成される。

【0026】格フラグ606aは欄の格情報の種類を入れる。属性ポインタ606bは属性リスト606cへのポインタを入れる。属性リスト606cは欄の概念の種類を入れる。リストポインタ606dは属性リストを複数個用いる場合の次属性リストへのポインタを入れる。

【0027】図1は、本実施例の知識ベース生成方法のフローチャートである。本実施例の知識ベース生成方法は処理102から処理116までの処理により実現される。以下、逐次、処理の説明を行う。

【0028】処理102はプログラム情報212を知識ベース218に変換する処理で、処理102の詳細を図7から図11までを用いて説明する。

【0029】処理102は図7の処理702から処理712までの処理に分けられる。

【0030】処理702はプログラム情報212のテーブル一覧308のデータをネームテーブル602および階層関係604に格納する処理である。これは「概念」という概念の下位概念としてテーブル名一覧308のデータを知識ベース218に格納するための処理である。

【0031】さらに、処理702の詳細を図8において説明する。処理702は、ネームテーブル602中のデータの格納位置を変数jに、階層関係604中のデータの格納位置を変数kに代入し(処理802)、テーブル一覧308のデータ件数のカウンタiに1を代入し(処理804)、テーブル一覧308のデータ件数を変数xに代入する(処理806)。

【0032】以下、 $i > x$ が成り立つまで(処理808)、ネームテーブル602のj行めの欄602bにテーブル一覧308の欄308aのデータを代入し(処理810)、ネームテーブル602のj行めの欄602aにj-1を代入し(処理812)、階層関係604のk行めの欄604aに0を代入し(処理814)、階層関係604のk行めの欄604bにj-1を代入し(処理816)、iとjとkにそれぞれ1を加える(処理818)処理を繰り返す。

【0033】処理704は「型」という概念を知識ベース218に作る処理である。これはあとでプログラム情報212の欄304bや欄306bのデータを知識ベースに格納する場合の上位概念となる。処理704は処理702においてテーブル一覧308の代りに「型」という1件のデータをネームテーブル602および階層関係604に格納するのと等価の処理である。

【0034】処理706は知識ベース218に「型」、

「含む」という名前の事実テーブルを新規に作る処理である。「型」という事実テーブルを図9の902に、

「含む」という事実テーブルを904に示す。ただし、これらのテーブル中のデータは処理706では作られない。

【0035】事実テーブル902は2個の欄を持つ。第1欄は格フラグとして「の」を持ち、属性リストとして「外部変数」と「自由変数」のIDを持つ。第2欄は格フラグとして「は」を持ち、属性リストとして「型」のIDを持つ。

【0036】事実テーブル904も2個の欄を持つ。第1欄は格フラグとして「は」を持ち、属性リストとして「サブルーチン」を持つ。第2欄は格フラグとして「を」を持ち、属性リストとして「自由変数」のIDを持つ。

【0037】概念IDはネームテーブル602の欄602aに格納されているので、属性リストのIDは欄602bをキーフィールド、「外部変数」、「自由変数」、「型」をキーとして欄602aから取り出したものを用いる。

【0038】処理706は事実テーブル902、904で示される形式のテーブルを作る。属性リスト中の値はネームテーブル602の欄602bの値が「サブルーチン」、「外部変数」、「自由変数」、「型」である欄602aの値を用いる。

【0039】処理708はプログラム情報212のサブルーチン302のデータをネームテーブル602および階層関係604に格納する処理である。これは、「サブルーチン」という概念の下位概念としてサブルーチン302のデータを知識ベース218に格納する処理である。

【0040】処理708は処理702において、テーブル一覧308の代りにサブルーチン302の欄302aのデータを格納対象とし、階層関係604の「上位」欄604aに代入するデータをネームテーブル602の欄602bの値が「サブルーチン」である欄602aの値を用いるように変更したものと等価である。

【0041】処理710はプログラム情報212の外部変数304のデータをネームテーブル602および階層関係604に格納し、処理706で作った事実テーブル902にデータを格納する処理である。これは、「外部変数」および「型」という概念の下位概念として外部変数304のデータを知識ベース218に格納し、外部変数の持つ型に関する事実を処理706で作った事実テーブル902に格納する処理である。

【0042】処理710の詳細は図10を用いて説明する。処理710では、カウンタiに1を代入し、外部変数304のデータ件数を変数aに代入する(処理1002)。次に、 $i > a$ の条件(処理1016)が成り立つまで、次の処理1004から1014の処理を繰り返す。

す。

【0043】処理1004は、処理702の要領で、外部変数304のi行めの欄304aのデータをネームテーブル602および階層関係604に格納する。このとき、階層関係604の「上位」欄604aの値は欄602bをキーフィールド、「外部変数」をキーとして欄602aから取り出した値を用いるようにする。

【0044】処理1006は、外部変数304のi行めの欄304bのデータを変数wに代入し、ネームテーブル602の欄602aの全てのデータを配列vに代入する。

【0045】配列vにwの値が含まれるという条件（処理1008）が成り立てば変数wに該当する概念のIDをネームテーブル602の欄602aからとりだし変数yに代入し（処理1009）、成り立たなければ、処理702の要領で、外部変数304のi行めの欄304bのデータをネームテーブル602および階層関係604に格納する。このとき、階層関係604の「上位」欄604aの値は欄602bをキーフィールド、「型」をキーとして欄602aから取り出した値を用いるようにする（処理1010）。

【0046】処理1012は、外部変数304のi行めのデータを事実テーブル902に追加する。このとき、事実テーブル902の欄902aに追加する値は処理1004でネームテーブル602に格納したデータのIDを用い、事実テーブル902の欄902bに追加する値は、処理1008が成功したときは処理1009の変数yを、失敗したときは処理1010でネームテーブル602に格納したデータのIDの値を用いる。処理1014は、カウンタiに1を加える。

【0047】処理712はプログラム情報212の自由変数306のデータをネームテーブル602および階層関係604に格納し、処理706で作った事実テーブル902,904にデータを格納する処理である。これは、「自由変数」および「型」という概念の下位概念として自由変数306のデータを知識ベース218に格納し、自由変数の持つ型に関する事実および自由変数の存在するサブマシンに関する事実を知識ベース218に格納する処理である。

【0048】処理712の詳細を図11を用いて説明する。処理712ではまず、カウンタiに1を代入し、自由変数306のデータ件数を変数aに代入する（処理1102）。次に、i>aの条件（処理1118）が成り立つまで、次の処理1104から1116の処理を繰り返す。

【0049】処理1104は、処理702の要領で、自由変数306のi行めの欄306aのデータをネームテーブル602および階層関係604に格納する。このとき、階層関係604の「上位」欄604aの値は欄602bをキーフィールド、「自由変数」をキーとして欄602aから取り出した値を用いるようにする。

【0050】処理1106は、自由変数306のi行めの欄306bのデータを変数wに代入し、ネームテーブル602の欄602aの全てのデータを配列vに代入する。

【0051】配列vにwの値が含まれるという条件（処理1108）が成り立てば変数wに該当する概念のIDをネームテーブル602の欄602aからとりだし変数yに代入し（処理1109）、成り立たなければ、処理702の要領で、自由変数306のi行めの欄306bのデータをネームテーブル602および階層関係604に格納する。このとき、階層関係604の「上位」欄604aの値は欄602bをキーフィールド、「型」をキーとして欄602aから取り出した値を用いるようにする（処理1110）。

【0052】処理1112は、自由変数306のi行めのデータを事実テーブル902に追加する。このとき、事実テーブル902の欄902aに追加する値は処理1104でネームテーブル602に格納したIDを用い、事実テーブル902の欄902bに追加する値は、処理1108が成功したときは処理1109でyに代入した値、失敗したときは処理1110でネームテーブル602に格納したIDを用いる。

【0053】処理1114は、自由変数306のi行めのデータを事実テーブル904に追加する。このとき、事実テーブル904の欄904aに追加する値は欄602bをキーフィールド、自由変数306のi行めの欄306cのデータをキーとして欄602aから取り出した値を用い、事実テーブル904の欄904bに追加する値は欄602bをキーフィールド、自由変数306のi行めの欄306aのデータをキーとして欄602aからとり出した値を用いる。

【0054】処理1116は、カウンタiに1を加える。

【0055】処理102を実行した結果の例を図9にあげる。

【0056】処理104はテンプレート214から事実テーブルを作成する処理である。処理104の詳細を図12から図15を用いて説明する。

【0057】処理104の詳細処理を図12を用いて説明する。処理104では、まず、変数jにテンプレート214のデータ件数を代入し、変数kにネームテーブル602のデータ格納位置を代入し、変数lに階層関係604のデータ格納位置を代入し、カウンタiに1を代入する（処理1202）。次に、i>jの条件（処理1216）が成り立つまで、次の処理1204から1214の処理を繰り返す。処理1204は、テンプレート214のi行めの欄214aのデータ名称を持つ事実テーブルを生成する。

【0058】処理1204の詳細を図13で説明する。処理1204では変数xにテンプレートのi行めの欄2

14aのデータを代入し(処理1301)、xという事実テーブルが知識ベース218に存在するか判定(処理1302)し、存在するならばxで示されるテーブルに新規に欄を1個追加し(処理1304)、存在しなければ欄を1個持つxという名前の事実テーブルを知識ベース218に新規に作る(処理1303)。

【0059】処理1206は、処理1304で作った欄の格フラグ606aにテンプレート214のi行めの欄214bのデータを代入する。

【0060】処理1208は、処理702の要領で、テンプレート214のi行めの欄214cのデータをネームテーブル602および階層関係604に追加する。

【0061】処理1208の詳細を図14で説明する。処理1208では変数yに概念「概念」のIDを代入し、変数mにネームテーブル602の格納位置を代入し、変数nに階層関係604の格納位置を代入(処理1402)し、ネームテーブル602の欄602bまたは欄602cのリストにテンプレート214のi行めの欄214cのデータが含まれているか調べ(処理1404)、含まれているか判定(処理1406)し、含まれているならば、ネームテーブル602のm行めの欄602bにテンプレート214のi行めの欄214cのデータを代入(処理1408)、ネームテーブル602のm行めの欄602cにm-1を代入し、階層関係604のn行めの欄604aにyを代入し、階層関係604のn行めの欄604bにm-1を代入する(処理1410)。

【0062】含まれていなければ、処理1208を終了する。

【0063】処理1210は、処理1204で作った欄の属性リストを1個作る。

【0064】処理1212は、処理1210で作った属性リストを図14の処理1410でネームテーブルの欄602aに代入した値を代入する。

【0065】処理1214は、カウンタiに1を加える。

【0066】処理104を実行した結果を図15に挙げる。

【0067】処理106は日本語の意味情報216の件数を変数saに代入し、カウンタsiに1を代入する。

【0068】処理108の条件 $s_i > sa$ が成立するまで処理110、112、114を繰り返す。

【0069】処理110はs_i件目の日本語の意味情報216のノード504中の語50aを知識ベース218のネームテーブル602と階層関係604に格納する。処理110の詳細を図16、図17を用いて説明する。

【0070】図16は処理110の詳細フローである。処理110は、配列xを空にし、動詞テーブル502の最初の格のポインタ502cを変数pに代入する(処理1602)。pが空(処理1604)になるまで処理1606から処理1622までを繰り返す。

【0071】pの指すノードの品詞を判定(処理1606)し、動詞でなければ配列xにpの指すノードの語502aを追加する(処理1608)。動詞ならば処理1612に飛ぶ。

【0072】処理1612は、pの指すノードの格名称フラグ504cが“=”か判定し、“=”ならばpの指すポインタ504dをpに代入し、処理1606に戻る。

【0073】“=”でなければ、処理1614に飛ぶ。

【0074】処理1614は、配列xのデータがネームテーブル602の欄602bまたは欄602cの同義語リストに1個でも含まれるかどうかネームテーブル602を逐次検索して判定する。

【0075】含まれるならば、ネームテーブル602で配列xのデータを含む行の番号をyに代入し(処理1615)、配列xのデータ数>1の判定(処理1616)を行う。処理1616の判定が成り立つならば配列xのデータをネームテーブル602のy行めの同義語リストにデータの重複が発生しないように追加する(処理1620)。

【0076】含まれないならば、配列xのデータをネームテーブル602と階層関係604に追加する(処理1618)。

【0077】処理1618の詳細を図17で説明する。処理1618では、動詞と格によって決まるIDを事実テーブルの属性リストからとりだし変数jに代入し、配列xのデータ数を変数jに代入し、ネームテーブル602のデータ格納位置を変数kに代入し、階層関係604のデータ格納位置を変数lに代入し、カウンタnに1を代入する(処理1702)。

【0078】処理1704は、ネームテーブル602のk行めの欄602bに配列xのn番めのデータを代入する。

【0079】処理1706は、ネームテーブル602のk行めの欄602aにk-1を代入する。

【0080】処理1708は、カウンタnに1を追加する。

【0081】処理1710は、 $n > j$ を判定し、 $n <= j$ ならば、ネームテーブル602のk行めに同義語リストを新規に追加(処理1712)し、同義語リストに配列xのn番めのデータを代入(処理1714)し、処理1708に戻る。

【0082】 $n > j$ ならば、階層関係604の1行めの欄604aにzを代入(処理1716)し、階層関係604の1行めの欄604bにk-1を代入する(処理1718)。

【0083】処理1622は処理1614が成功したときは処理1615でyに代入した行の欄602aのデータを、失敗したときは処理1706で欄602aに代入したデータを、pで示される格のID欄502dに代入する。

【0084】処理1622は、pに次の格のポインタ502cを代入する。

【0085】処理112はs i iめの日本語の意味情報216を処理104で作った事実テーブルに格納する。

【0086】処理112の詳細を図18を用いて説明する。処理112は、動詞テーブル502の語502aと同名の事実テーブルを探し(処理1802)、事実テーブルの存在を判定(処理1804)する。

【0087】事実テーブルが存在しなければ処理112を終了する。

【0088】事実テーブルが存在すれば最初の格のポインタ502cを変数xに代入し、処理1802で見つけた事実テーブルのデータ格納位置を変数iに代入する(処理1806)。

【0089】x=空(処理1808)が成立立つまで処理1810から処理1818までを繰り返す。

【0090】処理1810は、処理1802で見つけた事実テーブルから動詞テーブル502の格名称フラグ502bと同名の欄を探す。

【0091】処理1812は、処理1810で該当する欄が見つかったかの判定を行う。

【0092】見つかったならば、xで指される格のID欄502dのIDをaに代入(処理1814)し、処理1810で見つけた欄のi行めにaを代入する(処理1816)。

【0093】見つからなければ、処理1818に飛ぶ。

【0094】処理1818は、xの次の格のポインタをxに代入する(処理1818)。

【0095】処理110と処理112を実行したあとの知識ベースを図19に挙げる。

【0096】処理114は処理104で作った事実テーブルの属性リスト606cに事実テーブルに格納した概念の上位概念を追加する。

【0097】処理114の詳細を図20を用いて説明する。処理114は、事実テーブルの欄の数を変数jに代入し、カウンタiに1を代入する(処理2002)。i>j(処理2004)が成立立つまで処理2006から処理2018を繰り返す。

【0098】処理2006は、i番めの欄に新しく追加されたデータの上位概念のIDを概念階層604の欄604aからとりだし変数zに代入(処理2006)する。

【0099】処理2008は、i番めの欄の属性リストの先頭ポインタをpに代入する。

【0100】p=空(処理2010)の判定を行い、p=空ならばzを属性リストに追加(処理2012)し、処理2018に飛ぶ。

【0101】p=空でなければpの指す属性リストの値=zの判定を行い(処理2014)、等しくなければ、pの次の属性リストへのポインタをpに代入(処理201

6)し、処理2010に戻り、等しければ処理2018に飛ぶ。

【0102】処理2018は、カウンタiに1を加える。

【0103】欄属性の変更処理の実行後の知識ベースを図21に挙げる。

【0104】処理116は変数s iに1を加える。

【0105】本実施例の効果は、プログラム情報中の名称を表す知識を日本語の意味情報中に現れる語をつかって呼び出せる知識ベースを作ることができる。

【0106】さらに、プログラム情報212に位置情報を加えたものを知識ベースに変換する処理を処理102の代りに実行する実施例について図2から図26を用いて説明する。

【0107】位置情報を持つプログラム情報を図22に挙げる。図22のテーブル2202はテーブル名称として「サブルーチン」を持ち、サブルーチンの名称を欄2202aに持ち、サブルーチンの存在するファイル名を欄2202bに持ち、ファイル中におけるサブルーチンの先頭位置を2202cに持ち、ファイル中におけるサブルーチンの後尾位置を2202dに持つ。

【0108】次に、位置情報を格納する知識ベース218中の事実テーブルを図23に挙げる。図23のテーブル2302はテーブル名称が「位置する」であり、第1欄の格フラグ2302aに「は」を持ち、第2欄の格フラグ2302bに「で」を持ち、第3欄の格フラグ2302cに「から」を持ち、第4欄の格フラグに「まで」を持つ。

【0109】それぞれの欄の属性リスト2304には第1欄から「サブルーチン」、「ファイル名」、「数」、「数」という概念のIDが格納される。

【0110】処理102の代りに実行される位置情報を持つプログラム情報を知識ベースに変換する処理について図24を用いて説明する。

【0111】まず、処理702を実行し、「型」、「ファイル名」、「数」という名前をネームテーブル602および階層関係604に格納する(処理2404)。

【0112】処理2404は処理704において「型」、「ファイル名」、「数」という3件のデータを格納するのと等価の処理である。

【0113】処理2406は知識ベース218に「型」、「含む」、「位置する」という名前の事実テーブルを新規に作る処理である。

【0114】処理2408は位置情報を持つプログラム情報であるテーブル2202のデータを事実テーブル2302に格納する処理である。

【0115】処理2408の詳細を図25で説明する。処理2408では、カウンタiに1を代入し、サブルーチン2202のデータ件数を変数jに代入する(処理2502)。

【0116】次に、 $i > j$ の条件 (処理2504) が成り立つまで、次の処理2506から処理2524までを繰り返す。

【0117】処理2506は、サブルーチン2202の i 行めの欄2202aのデータをネームテーブル602および階層関係602に追加する。このとき、階層関係604の欄604aの値は欄602bをキーフィールド、「サブルーチン」をキーとして欄602aから取り出した値を用いる。

【0118】処理2508は、配列 w にネームテーブル602の欄602bおよび欄602cの同義語リストの全データを代入する。

【0119】配列 w がサブルーチン2202の i 行めの欄2202bのデータを含むか判定 (処理2510) し、含めば処理2514に飛び、含まなければサブルーチン2202の i 行めの欄2202bのデータをネームテーブル602および階層関係602に追加する (処理2512)。このとき、階層関係604の欄604aの値は欄602bをキーフィールド、「ファイル名」をキーとして欄602aから取り出した値を用いる。

【0120】配列 w がサブルーチン2202の i 行めの欄2202cのデータを含むか判定 (処理2514) し、含めば処理2518に飛び、含まなければサブルーチン2202の i 行めの欄2202cのデータをネームテーブル602および階層関係602に追加する (処理2518)。このとき、階層関係604の欄604aの値は欄602bをキーフィールド、「数」をキーとして欄602aから取り出した値を用いる。

【0121】配列 w がサブルーチン2202の i 行めの欄2202dのデータを含むか判定 (処理2518) し、含めば処理2522に飛び、含まなければサブルーチン2202の i 行めの欄2202dのデータをネームテーブル602および階層関係602に追加する (処理2520)。このとき、階層関係604の欄604aの値は欄602bをキーフィールド、「数」をキーとして欄602aから取り出した値を用いる。

【0122】処理2522は、処理2406で作った事実テーブル2302にサブルーチン2202のデータに対応するIDを格納する。

【0123】サブルーチン2202の欄2202aのデータに対応するIDは事実テーブル2302の欄2302aに格納し、サブルーチン2202の欄2202bのデータに対応するIDは事実テーブル2302の欄2302bに格納し、サブルーチン2202の欄2202cのデータに対応するIDは事実テーブル2302の欄2302cに格納し、サブルーチン2202の欄2202dのデータに対応するIDは事実テーブル2302の欄2302dに格納する。

【0124】IDは処理2506、処理2512、処理2516、処理2520でネームテーブル602に格納

したものをを用いる。

【0125】処理2524は、カウンタ i に1を加える。

【0126】位置情報を持つプログラム情報2202を知識ベース218に変換した結果を図26に挙げる。

【0127】本実施例の効果は、プログラム情報中の位置情報を日本語の意味情報中に現れる語をつかって呼び出せる知識ベースを作ることができる。

【0128】

【発明の効果】本発明によれば、知識ベースに格納する語が辞書にない語でも、コンピュータプログラム中で使用される言葉、または、コンピュータプログラム中使用される言葉と等価な意味合いを持つ言葉であることが日本語の意味構造からいえば、知識ベースに格納する語はコンピュータプログラム中で使用される言葉であるという語彙が決定できる。

【図面の簡単な説明】

【図1】知識ベース生成方法の処理フローチャート。

【図2】知識ベース生成方法を実現する時のブロック図。

【図3】プログラム情報の説明図。

【図4】テンプレートの説明図。

【図5】日本語の意味情報の説明図。

【図6】知識ベースの説明図。

【図7】プログラム情報を知識ベースに変換する処理フローチャート。

【図8】テーブル一覧のデータをネームテーブルと階層関係に格納する処理のフローチャート。

【図9】プログラム情報を格納後の知識ベースの説明図。

【図10】「外部変数」テーブルの情報を知識ベースに格納する処理フローチャート。

【図11】「自由変数」テーブルの情報を知識ベースに格納する処理フローチャート。

【図12】テンプレートから事実テーブルを作成する処理フローチャート。

【図13】事実テーブル作成処理フローチャート。

【図14】属性データをネームテーブルと階層関係に追加する処理フローチャート。

【図15】テンプレートから生成した事実テーブルの説明図。

【図16】日本語照合および概念登録の処理フローチャート。

【図17】日本語の意味情報中の語をネームテーブルと階層関係に追加する処理フローチャート。

【図18】日本語の意味情報を事実テーブルに格納する処理フローチャート。

【図19】日本語の意味情報格納後の知識ベースの説明図。

【図20】事実テーブルの欄属性の変更処理フローチャ

ート。

【図21】欄属性の変更結果の説明図。

【図22】位置情報を持つプログラム情報の説明図。

【図23】位置情報を格納する事実テーブルの説明図。

【図24】位置情報を持つプログラム情報を知識ベースに変換する処理フローチャート。

【図25】位置情報を持つ「サブルーチン」テーブルの情報を知識ベースに変換する処理フローチャート。

【図26】位置情報を持つプログラム情報を持つ知識ベ

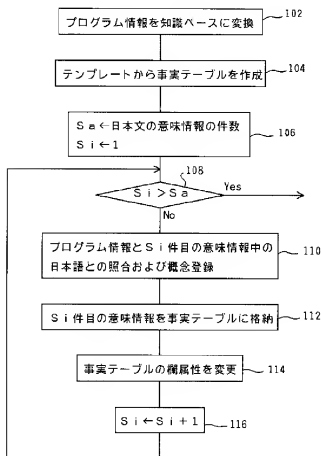
ースの説明図。

【符号の説明】

処理102…プログラム情報を知識ベースに変換する処理、処理104…テンプレートから事実テーブルを作成する処理、処理110…プログラム情報と意味情報中の日本語との照合および概念登録処理、処理112…意味情報を事実テーブルに格納する処理、処理114…事実テーブルの欄属性を変更する処理。

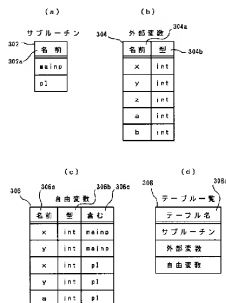
【図1】

図 1



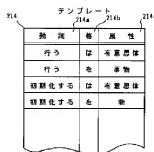
【図3】

図 3



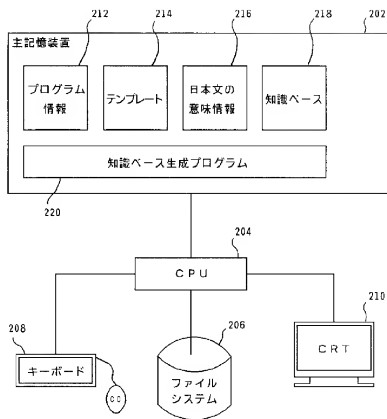
【図4】

図 4

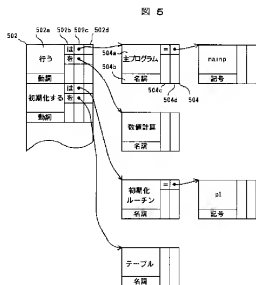


【図2】

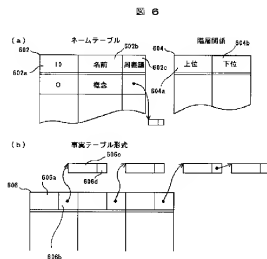
図 2



【図5】

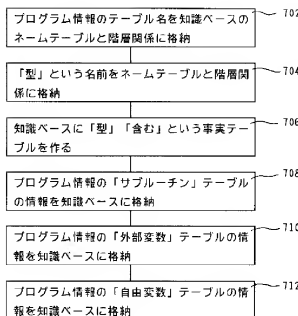


【図6】



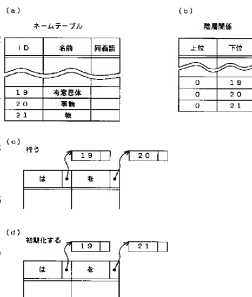
【図7】

図 7



【図15】

図 15

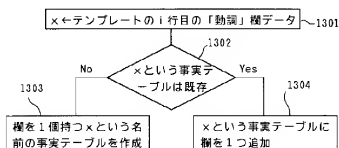


【図19】

図 19

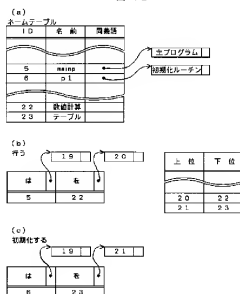
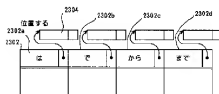
【図13】

図 13



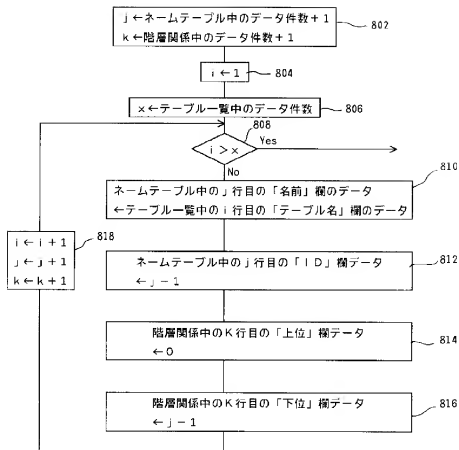
【図23】

図 23



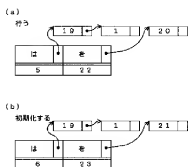
【図8】

図 8



【図21】

図 21



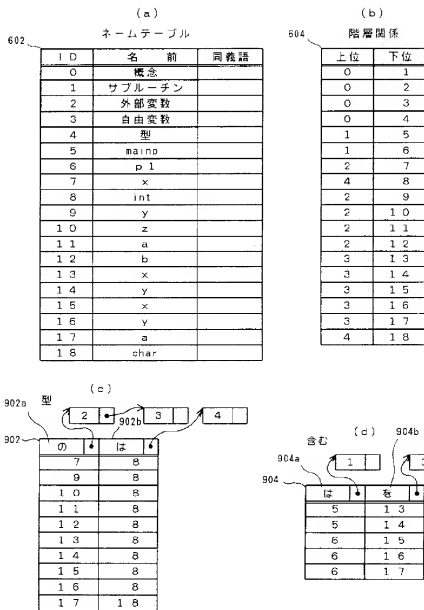
【図22】

図 22

サブルーチン			
名前	ファイル名	から	まで
mainp	f 1	5	5 0
P 1	f 1	5 2	8 0

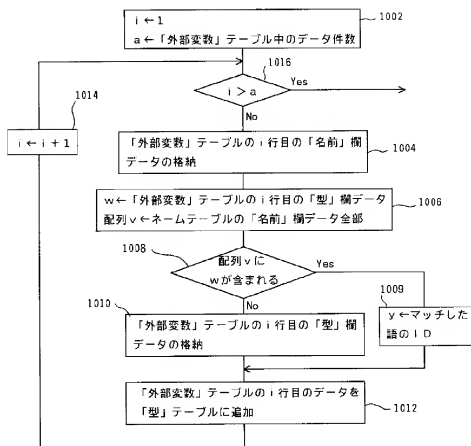
【図9】

図 9



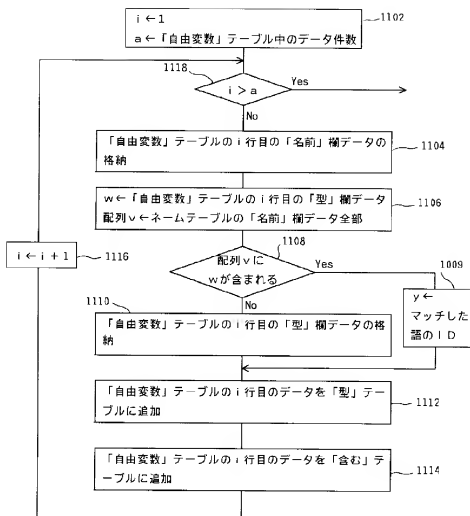
【図10】

図 10



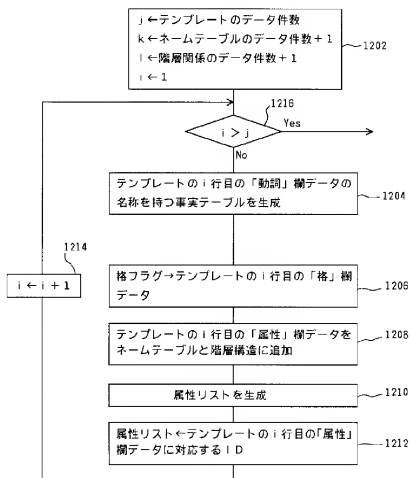
【図11】

図 11



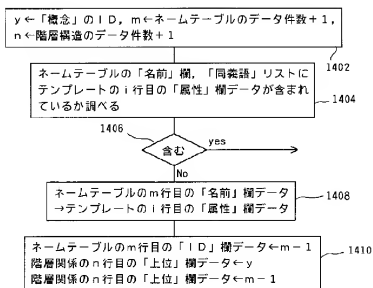
【図12】

図 12



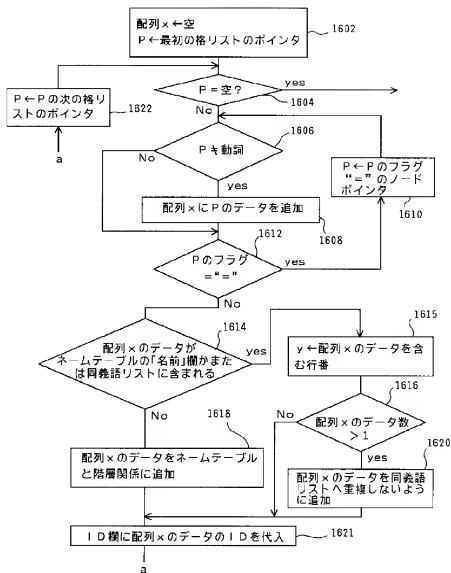
【図14】

図 14



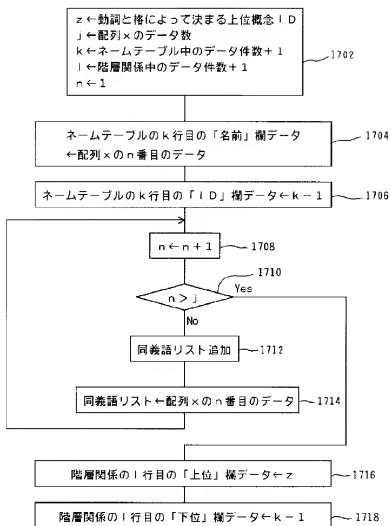
【図16】

図 16



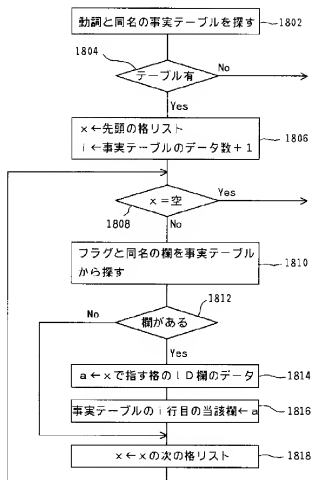
【図17】

図 17



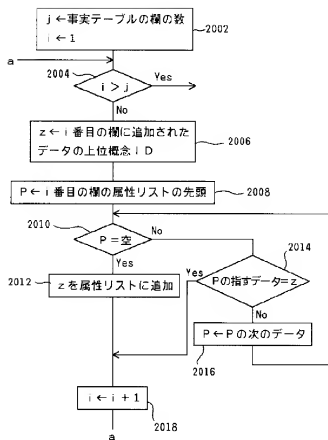
【図18】

図 18



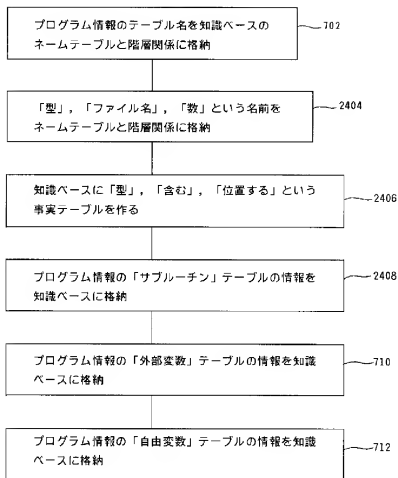
【図20】

図 20



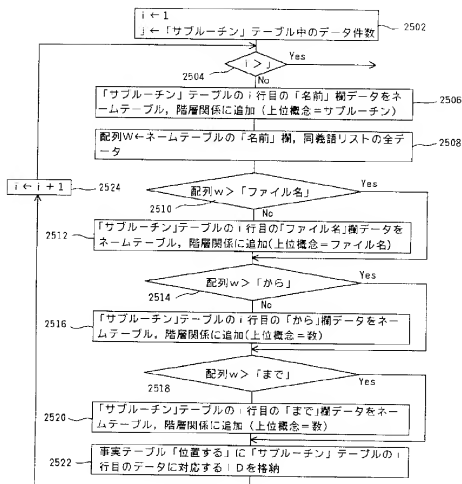
【図24】

図 24



【図25】

図 25



【図26】

図 26

(a)
ネームテーブル

ID	名前	同義語
0	概念	
1	サブルーチン	
2	外部変数	
3	自由変数	
4	型	
5	ファイル名	
6	数	
7	mainp	
8	f 1	
9	5	
10	5 0	
11	p 1	
12	5 2	
13	8 0	

(b)
階層関係

上位	下位
0	1
0	2
0	3
0	4
0	5
0	6
1	7
5	8
6	9
6	10
5	11
6	12
6	13

(c)
位置する

1	5	6	6
は	で	から	まで
7	8	9	10
11	8	12	13